European Journal of Human Genetics (2014), 1–9
© 2014 Macmillan Publishers Limited All rights reserved 1018-4813/14
www.nature.com/ejhg

ARTICLE

Regions of homozygosity identified by oligonucleotide SNP arrays: evaluating the incidence and clinical utility

Jia-Chi Wang[1], Leslie Ross[1], Loretta W Mahon[1], Renius Owen[1], Morteza Hemmat[1], Boris T Wang[1], Mohammed El Naggar[1], Kimberly A Kopita[1], Linda M Randolph[2], John M Chase[3], Maria J Matas Aguilera[4], Juan López Siles[4], Joseph A Church[5], Natalie Hauser[6], Joseph J Shen[6], Marilyn C Jones[7], Klaas J Wierenga[8], Zhijie Jiang[9], Mary Haddadin[1], Fatih Z Boyar[1], Arturo Anguiano[1], Charles M Strom[1] and Trilochan Sahoo*[,1,10]

Copy neutral segments with allelic homozygosity, also known as regions of homozygosity (ROHs), are frequently identified in cases interrogated by oligonucleotide single-nucleotide polymorphism (oligo-SNP) microarrays. Presence of ROHs may be because of parental relatedness, chromosomal recombination or rearrangements and provides important clues regarding ancestral homozygosity, consanguinity or uniparental disomy. In this study of 14 574 consecutive cases, 832 (6%) were found to harbor one or more ROHs over 10 Mb, of which 651 cases (78%) had multiple ROHs, likely because of identity by descent (IBD), and 181 cases (22%) with ROHs involving a single chromosome. Parental relatedness was predicted to be first degree or closer in 5%, second in 9% and third in 19%. Of the 181 cases, 19 had ROHs for a whole chromosome revealing uniparental isodisomy (isoUPD). In all, 25 cases had significant ROHs involving a single chromosome; 5 cases were molecularly confirmed to have a mixed iso- and heteroUPD15 and 1 case each with segmental UPD9pat and segmental UPD22mat; 17 cases were suspected to have a mixed iso- and heteroUPD including 2 cases with small supernumerary marker and 2 cases with mosaic trisomy. For chromosome 15, 12 (92%) of 13 molecularly studied cases had either Prader–Willi or Angelman syndrome. Autosomal recessive disorders were confirmed in seven of nine cases from eight families because of the finding of suspected gene within a ROH. This study demonstrates that ROHs are much more frequent than previously recognized and often reflect parental relatedness, ascertain autosomal recessive diseases or unravel UPD in many cases.
European Journal of Human Genetics advance online publication, 13 August 2014; doi:10.1038/ejhg.2014.153

INTRODUCTION

Oligonucleotide single-nucleotide polymorphism array (oligo-SNP) is an important diagnostic tool to identify both copy number variants (CNVs) and regions of homozygosity (ROHs) in patients with a wide variety of clinical indications. These ROHs are often unexpected and in many cases the clinical significance remains to be determined. Long contiguous ROHs may arise by a number of mechanisms. These include parental relatedness (or consanguinity) and chromosomal recombination or rearrangements. Depending on chromosomal distribution and cumulative extent, it may indicate background ancestral homozygosity, uniparental disomy (UPD) or parental consanguinity.[1–4]

Small random ROHs exist in all populations, including outbred cosmopolitan ones, and are believed to be a reflection of recombination rates and population history. A recent study using HapMap trios aimed at determining the prevalence of UPD in the general population suggested the existence of segmental UPD in 1 per 173 births (~0.6%) in the random population.[5] Such regions are usually <4 Mb as a contiguous stretch of the genome, particularly in outbred cosmopolitan populations according to a study of European populations.[6] In rare instances, ROH over 10 Mb as a single contiguous stretch has also been encountered and reported in certain individuals from different populations (Japan and Nigeria).[3] In addition to their clinical significance, the presence and extent of ROHs in a community or population mirrors the genetic structure and diversity of the population. It is now recognized that the effect of ROHs is not only significant for monogenic recessive disorders, but there is a role for recessive variants in complex genetic disorders.[7–9] When the regions with allelic homozygosity occur on multiple chromosomes, autozygosity resulting from inheritance of the genomic segment from parents who inherited it from a common ancestor is the most likely cause (a situation referred to as identity by descent (IBD)). In contrast, when such regions are restricted to a single chromosome, the possibility of UPD should be considered.

A recent clinical report of ~3000 cases studied by SNP array identified 59 (2%) cases with ROHs[1] that predicted the degree of

[1]Cytogenetics Laboratory, Quest Diagnostics Nichols Institute, San Juan Capistrano, CA, USA; [2]Division of Medical Genetics, Children's Hospital Los Angeles and Keck School of Medicine, University of Southern California, Los Angeles, CA, USA; [3]Division of General Pediatrics, Children's Hospital Los Angeles and Keck School of Medicine, University of Southern California, Los Angeles, CA, USA; [4]Molecular Biology Center GENETAQ, Málaga, Spain; [5]Division of Clinical Immunology and Allergy, Children's Hospital Los Angeles and Keck School of Medicine, University of Southern California, Los Angeles, CA, USA; [6]Children's Hospital Central California, Madera, CA, USA; [7]Rady Children's Hospital, San Diego, CA, USA; [8]Department of Pediatrics, Section of Genetics, The University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA; [9]Center for Computational Science, University of Miami, Miami, FL, USA
*Correspondence: Dr T Sahoo, Quest Diagnostics Nichols Institute, 33608 Ortega Highway, San Juan Capistrano, CA 92690, USA. Tel: +1 949 728 4336; Fax: +1 949 728 4979; E-mail: tsahoo001@gmail.com
[10]Current address: Combimatrix, Irvine, CA, USA.
Received 29 August 2013; revised 3 June 2014; accepted 10 July 2014

parental relatedness and aided the diagnosis of autosomal recessive diseases in two families; however, cases with UPD were not documented in this study. Another study of over 13 000 samples with developmental delay focused on the study of UPD only, with identification of 16 cases with isoUPD, 19 cases with hetero–isoUPD and 3 cases with segmental UPD.[4] There was no clinical study that provided an overall review of the finding of ROHs for all these three categories, that is, parental relatedness, autosomal recessive disorders and UPD, and this may be because of a huge variation for the threshold for reporting ROHs between laboratories,[10] and a lack of recommendations regarding the standards and guidelines for documenting suspected consanguinity.[11]

In this study, the largest and most comprehensive series reported so far, we report the finding of 832 (6%) cases with ROH from 805 families from 14 574 clinical cases referred to our laboratory: 651 (78%) cases with multiple ROHs and likely due to parental relatedness, and 181 (22%) cases with single-chromosome involvement. The data presented here represent the largest single collection of cases analyzed by oligo-SNP arrays leading to identification of a significant number of cases harboring clinically significant ROHs. This study further demonstrates that the identification of ROHs, in addition to CNVs, is much more frequent than previously recognized. Moreover, identification of ROHs is beginning to unravel clinically significant uniparental disomy and homozygosity for recessive disease gene mutations that would otherwise not have been suspected or identified.

MATERIALS AND METHODS

Patients

Patients with a broad range of clinical indications including intellectual disability, developmental delay, multiple congenital anomalies, dysmorphic features and pervasive developmental disorders were referred to our laboratory for oligo-SNP array studies. Over 85% of cases were referred for one or more of the above clinical indications. A small percentage of cases were studied by oligo-SNP arrays following normal conventional cytogenetic studies (karyotype and/or locus-specific FISH). The data were compiled from consecutive specimens that were referred to Nichols Institute at San Juan Capistrano over 2 years. The patients were majorly from general population in the United States, with <5% from Mexico and other countries.

Microarray analysis and homozygosity (HMZ) screening threshold setting (oligo-SNP)

Genomic DNA extracted from whole blood was utilized for analysis on either the Genome-Wide Human SNP Array 6.0 (40% of cases) or the CytoScan HD array (60% of cases) (Affymetrix, Santa Clara, CA, USA). The SNP array 6.0 contains 1.8 million genetic markers, including ~906 600 SNPs and 946 000 probes for the detection of copy number variation. The CytoScan HD array has more than 2.67 million probes, including 1.9 million nonpolymorphic copy number probes and 750 000 SNP probes. The density of SNP coverage between the two array platforms is similar; the mean SNP probe spacing is 3 kb and 3.6 kb for SNP 6.0 and Cytoscan HD, respectively. Genomic coordinates are based upon genome build 37/hg19 (2009). Hybridization, data extraction and analysis were performed as per the manufacturer's protocols. The Affymetrix Chromosome Analysis Suite (ChAS) Software version 1.0 or 2.0 was used for data analysis, review and reporting. The screening threshold used for length of ROH detection was set at 5 Mb; additional stringent parameters used to detect only true ROH included a minimum of 50 contiguous homozygous SNPs within a window, and <5 contiguous heterozygous SNPs (within a 100-kb interval). Regions with sparse SNP density were carefully evaluated to exclude false ROH calls. Copy number data for all ROH segments were carefully evaluated to rule out copy number losses (to rule out ROH due to hemizygosity). Although no consensus opinion exists regarding ROH size cutoffs, the choice of 5 Mb as the cutoff value is primarily derived from the

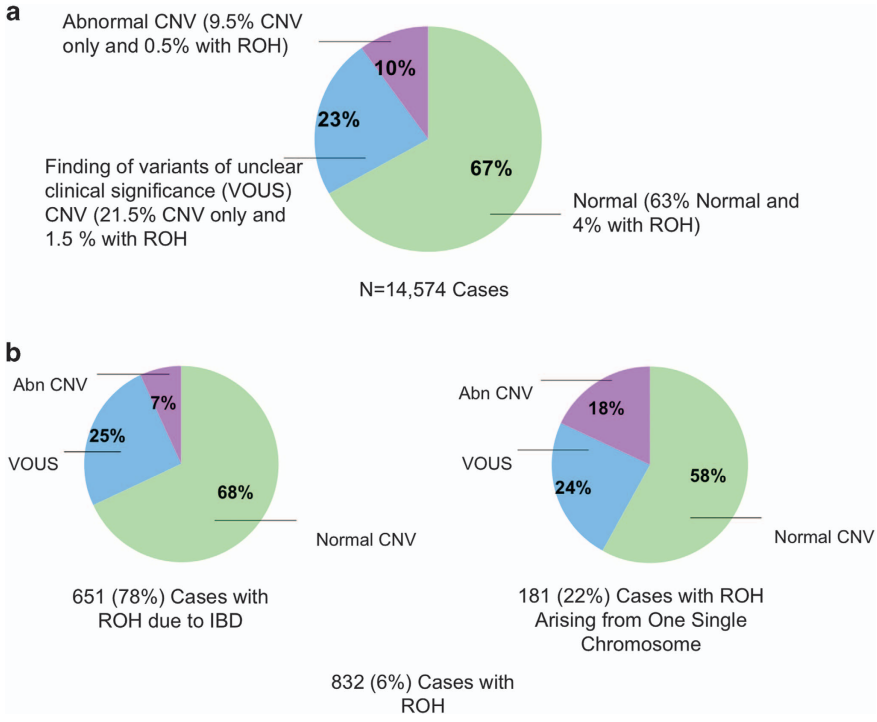

Figure 1 (**a**) A total of 14 574 cases were analyzed by oligo-SNP arrays: 9759 cases (67%) had a neutral copy number, 3362 cases (23%) had variants of unclear significance (VOUS) and 1453 cases (10%) had clinically significant copy number variations (CNVs). (**b**) In all, 832 cases (6%) had one or more reportable ROHs: 651 (78%) cases were interpreted as arising because of identity by descent (IBD), and 181 (22%) cases were suspected or confirmed as resulting from UPD. In cases due to IBD, 68% are copy number neutral, 25% had VOUS and 7% had clinically significant CNVs. In the cases with suspected or confirmed UPD, 58% are copy number neutral, 24% had VOUS and 18% had clinically significant CNVs.

study of European populations and recent data from other laboratories.[6,10] To avoid underestimation of the amount of ROH, the screening threshold was often varied so as not to miss ROHs just below the 5 Mb cutoff. ROH was reported as the presence of one segment of homozygosity >10 Mb or at least two ROHs >5 Mb in each region (with two or more segments on two or more chromosomes classified as IBD; and all segment(s) within one chromosome classified as possible UPD).

Quantifying ROHs and classification of degree of consanguinity

The amount of coverage of human genome in the oligo-SNP array was estimated by using the most telomeric probe for each chromosome on Cytoscan software and subtracting 180 Mb for the regions with heterochromatin around the centromeres; this was estimated to be 2700 Mb for autosomes. Except for the recurrent ROH segments, all other ROHs >5 Mb were included in the reporting. The classification of degree of consanguinity was based on the 95% confidence interval (CI) obtained using a methodology similar to the previous study from Sund *et al.*[1] CIs were calculated for five inbreeding percentages (25, 12.5, 6.25, 3.125 and 1.5625%) and for a total of 533 autosomal segments of 5 Mb each with SNP coverage.[1] Percentage homozygosity values (95% CI) obtained were very similar to that reported by Sund *et al*[1] (over 28.7% is indicative of parents being first-degree relatives or closer; 21.3–28.7% first; 15.3–21.3% first or second; 9.7–15.3% second; 8.3–9.7% second or third; 4.4–8.3% third; 4.2–4.4% third or fourth; 2.4–4.2% fourth; 2.0–2.4% fourth or fifth; 0.5–2.0% fifth).

Searching for potential candidate genes in ROHs for autosomal recessive disorders

Gene content and phenotype associations for ROH segments were evaluated by utilizing the web-based Genomic Oligoarray and SNP array evaluation tool, developed and made available by Wierenga *et al.*[12] By entering the genomic coordinates of all ROHs in this tool (www.ccs.miami.edu/ROH), a list of the OMIM genes mapping to those particular regions, and their associated (recessive) disorders, was generated for review.[12]

RESULTS

Recurrent ROHs

Five ROHs were recurrently identified in our patient population (and normal parents in many cases) and considered as polymorphic variants and not included in the reports and calculation: 3p21.31-p21.1 (chr3.hg19:g.47 586 513_53 002 150), 11p11.2-p11.12 (chr11.hg19:g.46 304 337_51 563 636), 16p11.2-p11.1 (chr16.hg19:g.29 194 414_35 220 544), Xp11.23-p11.22 (chrX.hg19:g.48 354 880_54 540 061) and Xq11.1q12 (chrX.hg19:g.61 932 503_67 123 772).

ROH due to IBD is a frequent finding in oligo-SNP array analysis

Of the 14 574 cases analyzed by oligo-SNP arrays, 9759 cases (67%) were neutral for copy number, 1453 cases (10%) had clinically significant CNVs and 3362 cases (23%) had copy number variants of unclear significance (VOUS) (Figure 1a). A total of 832 cases (6%) from 805 families had one or more reportable ROHs (Figure 1b): 548 cases with normal copy number, 79 cases with clinically significant CNVs and 205 cases with VOUS. These cases included sibling studies of a total of 50 patients from 23 families.

Of the 832 cases with ROHs, 651 (78%) cases had multiple ROHs from more than one chromosome and were interpreted as indicative of IBD. The cumulative ROHs ranged from 10.2 to 1096 Mb in size (mean = 135 Mb; median = 67 Mb; Figure 2). An estimate of the parental relatedness was determined: 31 (5%) cases with a first degree or closer kinship between the parents (see, Supplementary Figure 1); 58 (9%) cases second degree or closer; 130 (20%) cases third degree or closer (see, Supplementary Figure 2); 114 (17%) cases fourth degree or closer; 318 (49%) cases fifth degree or closer (Table 1a).

Sibling studies for 50 cases from 23 families (Table 1b) showed relatively concordant results in prediction of parental relatedness in 13 families, such as family 3 (three siblings each with ~876 Mb ROH or 32% HMZ in total) and family 20 with 289 Mb (10.7% HMZ) and

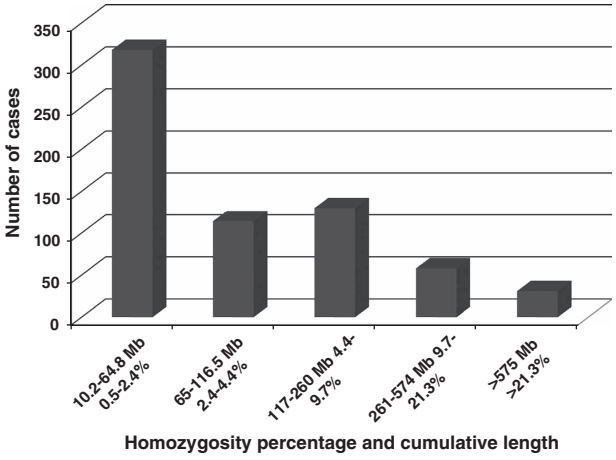

Figure 2 Total number of cases in each of the five inbreeding groups. The X axis represents the homozygosity percentage values and the corresponding cumulative length for the five groups (95% confidence interval). The Y axis represents the total number of cases in each group (Table 1a).

Table 1a Estimated degree of parental relatedness based on the cumulative ROH findings (*N* = 651)

Degree of relationship	*Theoretical percentage*	*Percent of homozygosity (95% CI; Sund et al*[1]*)*	*Percent of homozygosity (95% CI; current study)*	*Case number*	*Percentage*	*Added percentage*	*Percentage (Sund et al*[1]*)*
First or closer	>25%	>28.7%	>28.7%	13	2%		
First	25%	21.3–28.7%	21.3–28.7%	18	3%	5%	18%
First or second		15.3–21.3%	15.3–21.3%	9	1.4%		
Second	12.50%	9.7–15.3%	9.7–15.3%	49	7.5%	9%	17%
Second or third		8.3–9.7%	8.3–9.7%	19	3%		
Third	6.25%	4.6–8.3%	4.4–8.3%	111	17%	20%	30%
Third or fourth		4.2–4.6%	4.2–4.4%	5	0.8%		
Fourth	3.13%	2.6–4.2%	2.4–4.2%	109	16.5%	17%	15%
Fourth or fifth		1.6–2.6%	2.0–2.4%	41	6.3%		
Fifth	1.56%	0.5–1.6%	0.5–2.0%	277	42.5%	49%	20%
Total				651			

Abbreviation: CI, confidence interval at 95%.

Table 1b Sibling studies (23 families; *N* = 50)

Family	Age (years)	Sex	Clinical info	Chromosomes involved	Size (Mb)	% Of total genome 2700 Mb	Concordant
1	26	F	Branch chain amino acid disease	MULTIPLE	200	7.4%	No
	15	F	N/A	MULTIPLE	301	11.1%	
2	8	F	N/A	MULTIPLE	112	4.1%	No
	6	M	Mental retardation, behavior problem, speech delay, seizure disorder, right undescended testicle, parents are double first cousins	MULTIPLE	365	13.5%	
3	2	F	N/A	MULTIPLE	877	32.5%	Yes
	2	F	N/A	MULTIPLE	876	32.4%	
	2	F	N/A	MULTIPLE	876	32.4%	
4	15	F	N/A	MULTIPLE	65	2.4%	Yes
	5	M	N/A	MULTIPLE	36	1.3%	
5	3	M	N/A	MULTIPLE	53	2.0%	Yes
	2	F	N/A	MULTIPLE	78	2.9%	
6	10	M	Cleft lip and learning disabilities, speech delay, maternal aunt cerebral palsy	MULTIPLE	51	1.9%	Yes
	3	M	Speech delay; cleft lip; notable family history	MULTIPLE	48	1.8%	
7	6	M	Developmental delay; speech delay; autism spectrum disorder	MULTIPLE	28	1.0%	Yes
	4	M	N/A	MULTIPLE	13	0.5%	
8	5	F	N/A	MULTIPLE	27	1.0%	No
	6	M	N/A	1	10		
9	3	M	Mental retardation	22	10		No
	5	M	N/A	18	11		
	7	M	Mental retardation	MULTIPLE	39	1.4%	
10	11	F	Heart murmur, abnormal eye findings, nystagmus, pale fundus, depressed nasal bridge, open mouth, long fingers vision loss, light sensitivity	MULTIPLE	31	1.1%	No
	12	M	Heart murmur, prominent forehead, café-au-lait spots, abnormal eye findings, nystagmus, pale fundus, full lips, open mouth, cone dystrophy	MULTIPLE	395	14.6%	
11	2	M	N/A	MULTIPLE	52	1.9%	No
	7	F	N/A	17	21		
12	6	M	N/A	1	33		No
	11	M	Autism	MULTIPLE	26	1.0%	
13	5	M	developmental delay; autism spectrum disorder; macrocephaly	13	16		Yes
	7	F	developmental delay; autism spectrum disorder	13	16		
14	12	M	Developmental delay, speech delay, learning disability, brother is OSA12-11556 with segments of homozygosity	MULTIPLE	160	5.9%	No
	8	M	developmental delay; speech delay; learning disability; long bulbous tip nose	MULTIPLE	67	2.5%	
15	5	M	Autism	MULTIPLE	148	5.5%	Yes
	3	M	Speech delay	MULTIPLE	204	7.6%	
16	14	M	N/A	MULTIPLE	873	32.3%	No
	8	F	N/A	MULTIPLE	602	22.3%	
	10	F	N/A	MULTIPLE	612	22.7%	
17	6	F	N/A	MULTIPLE	194	7.2%	Yes
	3	F	N/A	MULTIPLE	186	6.9%	
18	5	M	N/A	MULTIPLE	39	1.4%	No
	17	F	N/A	MULTIPLE	427	15.8%	
	13	F	N/A	MULTIPLE	28	1.0%	
19	7	F	Developmental delay	MULTIPLE	61	2.3%	No
	8	F	Developmental delay	10	16		
20	3	M	Developmental delay, speech and motor delay, hypertonic, adducted thumbs, supernumerary nipple, hepatosplenomegaly, history of low platelets, parents are cousin	MULTIPLE	289	10.7%	Yes
	6	M	Intellectual disability, developmental delay, microcephaly, severe speech and motor delay, strabismus, pectus excavatum, prominent eyes, prenatal IUGR, parents are first cousin	MULTIPLE	329	12.2%	
21	7	M	N/A	MULTIPLE	31	1.1%	Yes
	15	F	N/A	MULTIPLE	26	1.0%	
22	10	M	N/A	MULTIPLE	46	1.7%	Yes
	1	M	Pyridoxamine 5′-phosphate oxidase deficiency	MULTIPLE	57	2.1%	
23	24	F	N/A	MULTIPLE	61	2.3%	Yes
	26	M	N/A	MULTIPLE	61	2.3%	

329 Mb ROHs (12.2% HMZ in two siblings). However, in some families, siblings may both have significant ROHs, but it did not predict the same degree of parental relatedness; for instance, in families 1 and 2, the percentage HMZ was 7.4–11.1% and 4.1–13.5%, respectively. Such discordant IBD predictions were detected in the siblings from families 1, 2, 8, 9, 10, 11, 12, 14, 16, 18 and 19.

Almost one-quarter of cases with ROHs arising from one single chromosome are because of suspected or confirmed isoUPD, mixed iso- and heteroUPD or segmental UPD

In 181 of 832 cases (22%), ROH from a single chromosome was detected. In 137 cases, the ROH was <25% of the size of the chromosome involved, raising the possibility of segmental UPD. Of the remaining 44 cases, 19 had ROH for a whole chromosome (uniparental isodisomy or isoUPD): 7 involving chromosome 15, 3 from chromosome 2, 2 from chromosome 7 and 1 each from chromosomes 1, 5, 8, 9, 16, 19 and 20 (Table 2a) (see, eg, Supplementary Figure 3). One case with UPD5 was associated with a small, mosaic supernumerary marker chromosome 5 (1.3 Mb in size) (18 of 20 cells), and one case with UPD8 and a familial nonmosaic marker chromosome 8 (6.1 Mb). In the remaining 25 cases, ROHs did not involve the whole chromosome (Table 2b) and included 8 from chromosome 15 (see, eg, Supplementary Figure 4) and 4 from chromosome 9. Concurrent ROHs and mosaic trisomy for chromosomes 9 and 18 were seen in 2 cases, respectively, and could most likely be because of trisomy rescue.[13] Concurrent ROHs (mixed iso- and heteroUPD) and small supernumerary marker chromosomes were seen in two cases, from chromosomes 1 and 4, respectively.[14]

Table 2a Cases with ROHs only involving a single chromosome

	Cases with segmental ROH from a single chromosome					
Chromosome	*Total cases*	*Cases*	*Smallest (Mb)*	*Largest (Mb)*	*Median (Mb)*	*Cases with ROH involving the whole chromosome*
1	14	13	10	110	13	1
2	17	14	10	27	13	3
3	13	13	11	27	13	0
4	8	8	10	42	19	0
5[a]	9	8	10	21	16	1
6	9	9	10	22	14	0
7	14	12	10	36	18	2
8[a]	6	5	11	17	14	1
9	7	6	13	78	29	1
10	6	6	10	19	13	0
11	13	13	11	31	15	0
12	9	9	10	25	11	0
13	5	5	10	19	16	0
14	5	5	10	33	17	0
15	15	8	13	47	26	7
16	5	4	12	30	21	1
17	3	3	10	21	18	0
18	6	6	10	35	17	0
19	4	3	10	30	20	1
20	2	1	21	62	42	1
21	1	1	11	11	11	0
22	3	3	10	21	19	0
X	7	7	11	74	23	0

[a]Positive for supernumerary marker chromosome (one case from each chromosome).

The majority of cases with ROHs from chromosome 15 are due to UPD15

Among the 15 cases with ROHs from chromosome 15 alone (Table 2c), 7 had isoUPD for the entire chromosome 15 and 8 had ROH from 13 to 47 Mb (Table 2c, Supplementary Figure 4). All 7 cases with ROH of whole chromosome 15, and 5 of 8 cases (63%) cases with segmental ROH 15 were molecularly confirmed to be clinically PWS or AS. Please note that both cases with ROH of only 13 and 16 Mb (16 and 20% of the size of whole chromosome 15) had Prader–Willi syndrome (Table 2b, cases 9 and 10; Table 2c, cases 10 and 11).

UPD11 in cases with ROHs from chromosome 11 alone

Among the 13 cases with ROHs from chromosome 11 alone, none involve the whole chromosome. One case with a clinical indication of hemihypertrophy and hypoglycemia in the neonatal period was revealed to have a segmental ROH in a mosaic state that was ~15 Mb in size (11p15.2-pter; arr[hg19] 11p15.5p15.2(198 510–15 029 576) × 2 mos hmz; Supplementary Figure 5). This segment includes the imprinted domain at 11p15.5 associated with Beckwith–Wiedemann syndrome (BWS, OMIM #130650). Methylation-sensitive multiplex ligation-dependent probe amplification (MLPA) performed elsewhere showed hypermethylation of IC1 (H19) and hypomethylation of IC2 (LIT1), consistent with mosaic paternal UPD11 and a clinical diagnosis of BWS.

Autosomal recessive disorders unmasked by ROHs

In nine patients (from eight families), identification of significant ROH followed by analysis of ROH segments for disease-causing genes helped molecular diagnosis in seven (Table 3). (Cases 1 and 2). The 3-year-old twin boy and girl with short stature and hydrocephalus, both with little response to growth hormone, were diagnosed to have Laron dwarfism. A homozygous c.181C>T transition, predicted to lead to a premature stop codon in the growth hormone receptor gene *GHR*, NM_000163.4:c.181C>T [p.(Arg61*)], was confirmed (OMIM #262500; growth hormone receptor, *GHR* at 5p13.1-p12). (3) A 3-year-old boy with primary microcephaly (FOC 43.5 cm at 3 years of age), developmental delay, speech delay, close-spaced eyes, epicanthal folds, downslanting palpebral fissures, large ears and smooth philtrum. A homozygous deletion of ~250 kb was found at 8p23.2-p23.1 and within a ROH, chr8.hg19:g.(6060654_6061169)_(6310738_6317266)del. The deleted region includes the microcephalin gene, *MCPH1*, and most likely resulting in primary autosomal recessive microcephaly 1 (OMIM #251200, *MCPH1* at 8p23.2) (Hemmat *et al.*, manuscript in preparation, 2014). (4) An 18-day-old male baby with organic aciduria and elevated C5OH was diagnosed with 3-methylglutaconic aciduria, type I (OMIM #250950; AU RNA binding protein/enoyl-CoA hydratase (*AUH*), at 9q22.31). A homozygous missense transition mutation (NM_001698.2:c.373C>T, [p.(Arg125Trp)]) was identified in the AU-specific RNA binding protein (*AUH*) gene, encoding a mitochondrial protein.[12] This sequence change affects a highly conserved amino acid residue located in a functional domain of the AUH protein. The homozygous c.373C>T change appears to be deleterious using several *in silico* prediction tools (SIFT, Polyphen-2, MutationTaster) and in the context of this patient's clinical presentation it is interpreted as most likely pathogenic. (5) A 23-day-old baby male with coarse features, unusual palate, pectus excavatum, long fingers, contractures and camptodactyly was

Table 2b Cases with one or more ROHs arising from a single chromosome (>25% of chromosome length, except for chromosome 15)

Case	Chromosome	ISCN	Total segments	Overall size	Ratio of ROH to whole chromosome	Associated abnormalities	Suspected or confirmed UPD
1	1	arr[hg19]1p22.1p11.2 (94 660 325–121 339 317) × 2 hmz, 1q31.1q32.1 (189 547 500–205 510 242) × 2 hmz, 1q42.12q44 (226 661 378–249 198 692) × 2 hmz	3	110	44%	Supernumerary marker chromosome	To be confirmed
2	4	arr[hg19] 4p13q12 (44 452 759–57 264 239) × 3, 4p16.1p13 (8 023 483–44 438 561) hmz, 4q22.3q28.1(96 302 989-127 376 679), hmz 4q32.2q35.1 (162 602 645–183 584 885) hmz, 47,XY, + mar.ish der(4)(D4Z1 +)[3], nuc ish der(4)(D4Z1 +)[26/76]	3	89	47%	Supernumerary marker chromosome	To be confirmed
3	9	arr[hg19] 9p21.3q22.33(23 642 676–101 084 087) × 2 hmz	1	78	55%	None	To be confirmed
4	9	arr[hg19] 9q21.11q21.33 (70 984 588–86 956 612) × 4, 9q21.33q34.3(86 956 612–140 955 352) × 2 hmz	1	54	38%	Triplication of 16 Mb immediately proximal to the ROH	Segmental UPD9pat
5	9	arr[hg19] 9p24.3q34.3 (192 129–141 025 328) × 2-3, 9q21.13q33.1(79 068 626–122 450 340) hmz	1	43	31%	Mosaic trisomy 9 in 11% of cells	To be confirmed
6	9	arr[hg19] 9q21.11q34.3(71 019 572–138 853 385) hmz	1	70	50%	None	To be confirmed
7	13	arr[hg19] 8p22 (13 356 543–14 659 874) × 4, 13q12.2q21.2(28 106 155–61 295 506) hmz, 13q22.1q33.1(74 941 789–104 625 030) hmz	2	63	66%	None	To be confirmed
8	14	arr[hg19] Xq12(65 624 620–66 171 491) × 3, 3q28(191 912 869–192 043 526) × 1, 14q13.2q24.1(35 878 430–68 197 350) × 2 hmz	1	33	38%	None	To be confirmed
9	15	arr 15q26.1q26.3(89 356 831–102 165 080) × 2 hmz	1	13	16%	None	Iso-heteroUPD15mat
10	15	arr 15q15.1q21.3(42 413 212–58 822 785) × 2 hmz	1	16	20%	None	Iso-heteroUPD15mat
11	15	arr 15q21.3q24.1(58 176 512–73 848 422) × 2 hmz	1	16	20%	None	Negative for PWS and AS
12	15	arr 15q21.3q24.1(54 014 881–74 573 931) × 2 hmz	1	21	26%	None	To be confirmed
13	15	arr[hg19] 15q21.3q26.1(58 474 697–89 452 672)x2 hmz	1	31	39%	None	Iso-heteroUPD15mat
14	15	arr[hg19] 15q14q23(36 597 898–68 726 159) × 2 hmz	1	32	40%	None	To be confirmed
15	15	arr 15q21.3q26.3(58 666 471–102 429 049) × 2 hmz	1	44	55%	None	Iso-heteroUPD15pat
16	15	arr[hg19] 1p31.1(74 010 837-74 849 251) × 3, 2q13(112 668 524–112 886 200) × 3, 15q11.2q13.3(22 752 398–31 590 959) × 2 hmz,15q21.1q25.3(49 231 110–87 298 245) × 2 hmz	2	47	59%	None	Iso-heteroUPD15mat
17	16	arr[hg19] 16p12.3q12.1(20 948 751–51 149 011) × 2 hmz	1	30	34%	None	To be confirmed
18	16	arr[hg19] 16p13.3p12.3(86 670-21 033 688) × 2 hmz, 16q23.1q24.1(78 969 980–84 414 164) × 2 hmz	2	26	29%	None	To be confirmed
19	18	arr[hg19] 18q11.1q21.31(21 653 163–56 219 590) × 2 hmz	1	35	44%	None	To be confirmed
20	18	arr[hg19] 18q11.2q21.1(21 860 416–45 546 224) × 2 hmz	1	24	30%	None	To be confirmed
21	18	arr[hg19] 18p11.32q23 (136 226–77 990 800) × 2-3, arr 18p11.31q12.1 (6 646 852–27 007 109) × 2-3 hmz, 47,XY, + 18[2]/46,XY[7].ish 18cen(D18Z1 × 3)[38]/ 18cen(D18Z1 × 2)[62]	1	21	27%	Mosaic trisomy 18 in 38% of cells	To be confirmed
22	19	arr[hg19] 16p13.11 (15 498 806–15 618 101) × 4, 19p13.12q13.32 (16 161 026–46 530 250) × 2 hmz	1	30	51%	None	To be confirmed
23	20	arr[hg19] 20p11.21q13.31(24 108 772–55 369 127) × 2 hmz	1	21	33%	None	To be confirmed
24	22	arr[hg19] 22q12.1q12.2(27 781 546–29 782 433) × 4, 22q12.2q13.33(29 793 640–51 010 112) × 2 hmz, nuc ish 22q12.1q12.2(RP11-419L7 × 3,CTD-2010F20 × 3)	1	21	62%	Triplication of 2 Mb immediately proximal to the ROH	Segmental UPD22mat
25	X	arr[hg19] 15q11.2(22 770 421–23 676 512) × 1, Xp21.2q21.32 (31 198 676-92 510 360) × 2 hmz	1	61	40%	None	To be confirmed

suspected to have mucolipidosis type II (OMIM #252500) and strongly suspected to harbor homozygous mutation in the gene *N-acetylglucosamine-1-phosphate transferase, alpha/beta subunits* (*GNPTAB*, at 12q23.2). (6) A 6-month-old male child with partial seizures, mild white matter hypoplasia with corpus callosum and enlargement of the lateral ventricles shown in brain MRI, abnormal

Table 2c Cases with ROH arising from chromosome 15

UPD15	*Size of ROH*	*Methylation study*
1	Whole	PWS positive
2	Whole	PWS positive
3	Whole	AS positive
4	Whole	AS positive
5	Whole	AS positive
6	Whole	AS positive
7	Whole	AS positive
8	47	PWS positive
9	31	PWS positive
10	13	PWS positive
11	16	PWS positive
12	44	AS positive
13	16	Negative for PWS or AS
14	32	Not confirmed
15	20	Not confirmed

Abbreviations: AS, Angelman syndrome; PWS, Prader–Willi syndrome.

EEG, hypotonia, wide inner canthal distance, widely spaced nipples and wide nasal bridge was suspected to have bare lymphocyte syndrome type II (OMIM #209920, *regulatory factor X-associated protein* (*RFXAP*) at 13q13.3). A homozygous nonsense mutation in exon 1 of the *RFXAP* gene predicted to result in a premature stop codon was identified (NM_000538.3:c.323T>A, [p.(Leu108*)]). (7) A 7-year-old female suspected to have Bardet–Biedl syndrome 4 (OMIM #209900, *BBS4* gene at 15q23). (8 and 9). An additional two cases with a confirmed recessive disorder with one having segmental UPD9 and the second with whole-chromosome UPD16, respectively (Table 3, cases 8 and 9). Case 8 is a 5-year-old girl with elevated citrullin, spastic quadriparesis, developmental delay and found to have autosomal recessive citrullinemia (OMIM #215700, *argininosuccinate synthetase 1* (*ASS1*) gene at 9q34.11) harbored a complex abnormality involving 9q including homozygosity for the 9q21.33-qter segment. A homozygous missense mutation in the *ASS1* gene (NM_000050.4:c.571G>A, [p.(Glu191Lys)]) was identified and found to have been inherited from a heterozygous carrier parent (case 4, Table 2b). This mutation has been previously described in at least two patients with autosomal recessive citrullinemia.[15,16] Case 9 is a 24-day-old female baby with whole-chromosome UPD16 suspected to have Bardet–-Biedl syndrome 2 (OMIM #209900 and 606151, *BBS2* gene at 16q12.2). A homozygous frameshift mutation in exon 14 of *BBS2* gene was identified and predicted to lead to a premature stop codon (NM_031885.3:c.1770delT, [p.(Phe590Leufs*8]). One parent of the proband was identified to be a heterozygous carrier for the identical mutation.

DISCUSSION

Implementation of oligo-SNP arrays has enabled simultaneous interrogation of genomic DNA for copy number variations and copy neutral regions of allelic homozygosity. The availability of this information proves invaluable in providing diagnostic information critical to the clinical evaluation of patients with constitutional genetic disorders. The cumulative extent of ROHs in any given case is determined by multiple factors such as parental relatedness, ethnicity, recombination events and chromosomal aberrations. The significance of such findings and deriving clues towards molecular testing and genotype–phenotype correlations necessitates critical evaluation of the array data in each individual case.

Currently, there is no consensus regarding the threshold for identification of runs or ROHs in constitutional cases.[10] The choice of 5 Mb as the cutoff value for identification of ROHs is primarily derived from study of European populations, demonstrating that outbred individuals rarely carried ROHs over 4 Mb in size.[6] However, the cutoff value for other populations is likely to be different.[17] In this study, 832 (6% of 14 574) cases were identified to harbor one or more ROHs, and this is much more frequent than previously reported.[1] It may be because of a lower but appropriate size threshold, a larger case series or sampling of populations in this study.

The classification of degree of consanguinity for 651 cases was based on the 95% CI calculated following a methodology similar to that used in a previous small study of 59 patients.[1] The significant agreement in the predicted percentage homozygosity for the five degrees of inbreeding (consanguinity) provides strong credibility to the methodology used and, additionally, lays the basis for reaching a consensus guideline for evaluation, interpretation and reporting of ROH. It is to be noted that, despite agreement in homozygosity percentage intervals for calculating parental relatedness, the frequency of cases for each of the five degrees was somewhat different. This can be attributed to a very large number of cases in this study that would reduce the possibility of a chance occurrence. Our data set includes close to 11 times more cases with IBD (651/59), thus strengthening our predictions. There is remote likelihood that the calculated frequencies may not reflect the true degree of consanguinity for each particular family as it has been demonstrated that the frequency of consanguinity is diverse among populations, thus leading to a varied size of extended tracts of homozygosity.[3,17]

In addition to providing an estimate for parental relatedness in cases with IBD, one surprising revelation was a significant number of cases ($N=31$; 5%) wherein parental relatedness could be assessed to be of the first degree or closer. This level of cumulative homozygosity of 25% of the genome or greater adds a new dimension to the clinical evaluation and counseling process. It is increasingly becoming a challenge about how to respond to and counsel for such results, and a need for genetics professionals to have a guideline has been requested.[18] The ethical, legal and social outcomes that these cases present and the approaches to dealing with the same are still evolving.[18,19]

In the sibling studies, the result showed only 13 of 23 families (~57%) had a concordant result in prediction of parental relatedness. Although in some families such as family 3 (Table 1b), three siblings had almost exact amount of ROHs (~876 Mb), for other families, ROH from one sibling alone will not reliably predict the degree of relatedness between parents. For instance, in family 10, both siblings had a very similar clinical phenotype, but the ROHs for each sibling was discrete (31 *vs* 395 Mb), thus precluding easy genotype–phenotype correlation, or raising the possibility of nonpaternity.

Another aspect of this study is the frequency of identification of UPD (in more than 40 cases) including isoUPD (19 cases), segmental UPD (2 cases) and iso- and heteroUPD (22 cases). The major drawback to having the inclusion criteria set for at least two ROHs on two separate chromosomes, as suggested by Sund *et al*,[1] is likely to increase the risk of missing single-chromosomal ROH that may be due to combined iso- and heteroUPD. We demonstrated that ROHs only from a single chromosome is frequently encountered and can involve any chromosome (Table 2a). Large chromosomes such as chromosomes 1–3 and chromosomes with imprinted genes are more frequently identified than other chromosomes.

Table 3 Findings in cases with confirmed autosomal recessive disorders

Case	*Age*	*Sex*	*Result*	*Interpretation*	*Chromosomes involved*	*ROH size (Mb)*	*AR disorders*	*OMIM*	*Gene*	*Chr band*	*Coordinates for relevant HMZ segment [hg19]*	*Mutation*
1	3 Years	F	No CNV, multiple ROH	IBD	Multiple	101	Laron dwarfism	262500	*GHR*	5p13.1-p12	5p15.1p12 (17 228 907–43 820 935) × 2 hmz	NM_000163.4:c.181C>T [p.(Arg61*)]
2 (twin sibling of case 1)	3 Years	M	No CNV, multiple ROH	IBD	Multiple	84	Laron dwarfism	262500	*GHR*	5p13.1-p12	5p15.1p12 (17 228 907–43 820 935) hmz	NM_000163.4:c.181C>T [p.(Arg61*)]
3	3 Years	M	No CNV, multiple ROH	IBD	Multiple	147–167	Primary, autosomal recessive microcephaly 1	251200	*MCPH1*	8p23.2	8p23.1p11.1 (8 107 313–43 778 914) × x2 hmz	chr8.hg19:g.(6060654_6061169)_(6310738_6317266)del
4	18 Days	M	Copy number gains at Xq28 and 10q26.11-q26.12, copy number loss at 16q23.1, multiple ROH	IBD	Multiple	284	3-methylglutaconic aciduria, type I	250950	*AUH* (nuclear gene encoding mitochondrial protein)	9q22.31	9q22.2q31.1 (93 622 149–107 626 389) × 2 hmz	NM_001698.2:c.373C>T, [p.(Arg125Trp)]
5	23 Days	M	One copy number gain at 15q13.3 with multiple ROH	IBD	Multiple	309.6	Mucolipidosis II alpha/beta	252500	*GNPTA, GNPTAB*	12q23.2	12q23.2q23.3(101 918 980–108 990 019) × 2 hmz	Not confirmed
6	6 Months	M	One copy number loss at 5p15.2 with multiple ROH	IBD	Multiple	30	Bare lymphocyte syndrome, type II, complementation group D	209920	RFXAP	13q13.3	13q13.2q14.13 (35 347 597–46 817 366) × 2 hmz	NM_000538.3:c.323T>A, [p.(Leu108*)]
7	7 Years	F	No CNV, multiple ROH	IBD	Multiple	15.4	Bardet-Biedl syndrome 4	209900	*BBS4*	15q23	15q23q25.1 (71 062 832–81 339 871) × 2 hmz	Not confirmed
8	5 Years	F	Triplication of 9q21.11-q21.33 with ROH on chromosome 9 only	UPD	9	54	Citrullinemia	215700	*ASS1*	9q34.11	9q21.33q34.3(86 956 612–140 955 352) × 2 hmz	NM_000050.4:c.571G>A, [p.(Glu191Lys)]
9	24 Days	F	UPD16	UPD	16	WHOLE	Bardet-Biedl syndrome 2	209900	*BBS2*	16q12.2	16q11.2q24.3(46 464 488–90 163 275) × 2 hmz	NM_031885.3:c.1770delT, [p.(Phe590Leufs*8)]

Abbreviations: ASS1, argininosuccinate synthetase 1; AUH, AU-specific RNA-binding protein; BBS2, Bardet-Biedl syndrome-2; BBS4, Bardet-Biedl syndrome-4; GHR, growth hormone receptor; GNPTA, N-acetylglucosamine-1-phosphotransferase; MCPH1, microcephalin; RFXAP, regulatory factor X-associated protein.

The ratio of both pure iso- and iso/heteroUPD15 in cases with ROH restricted to chromosome 15 is much higher than the other chromosomes (Tables 2b and c). In the cases with ROHs from chromosome 15, follow-up molecular studies (in 13 of 15 cases) confirmed UPD15 in 12 of 13 cases and helped reach a clinical diagnosis of PWS or AS. Two cases with small ROHs (<25% of the whole chromosome) were molecularly confirmed to have PWS (Table 2b, cases 9 and 10; Table 2c, cases 10 and 11). Therefore, the size of ROH on chromosome 15 cannot be used as a reliable predictor of whether the patient had UPD15 or not. This underlines the importance of follow-up analysis for these patients even if the ROH may not be large. The HMZ segments in cases with mixed iso- and heteroUPD did not show a significantly greater concentration of HMZ segments at the subtelomeric or terminal regions. This is somewhat surprising and contrary to theoretical expectations that single-chromosome ROHs reflective of UPD would be more frequent at subtelomeric regions and signify true illegitimate meiotic recombination.[4] There does not seem to be an easy explanation for this somewhat discrepant observation and our findings need further support from additional large studies. Future studies complemented by molecular confirmation of suspected UPD cases will be valuable. As an example, a case with 13 Mb of ROH (case 10, Table 2c) was located at 15q26.1-q26.3 (chr15:89 356 831–102 165 080) that is telomeric; another with 16 Mb of ROH (case 11, Table 2c) at 15q15.1-q21.3 (chr15:42 413 212–58 822 785) was not telomeric. We also demonstrated that parental origin of UPD15 can be identified by SNP array genotyping analysis if parental array data are available (Table 2c, case 8). The proband had two regions of homozygosity from chromosome 15 (a total of 47 Mb). A comparison of the genotype data between the proband and the mother disclosed both of the chromosomes 15 were inherited from the mother (UPD15mat). Molecular testing confirmed the diagnosis of PWS. Therefore, availability of parental genotype data allows imputation of parent of origin and may avoid the need for additional molecular studies in some cases.

For other chromosomes, it is likely that a significant number of these cases with ROH <25% of the whole chromosome in size and restricted to a single chromosome (although raising the possibility of segmental or whole chromosome UPD) may be coincidental in nature and molecular testing may exclude UPD (for example, case 13 in Table 2c). However, until a substantial number of such cases are actually evaluated by molecular analysis, no conclusion may be drawn.

Cytogenetic rearrangements have been widely known to cause UPD.[13] The concurrent presence of marker chromosome or mosaic trisomy and single-chromosomal ROH aids in the diagnosis of UPD.[13,14] Our finding also strongly supports the use of oligo-SNP arrays in the evaluation of marker chromosomes (even when mosaic) and mosaic trisomy as it may provide crucial evidence not only regarding the identity of the marker chromosome and trisomy, but also help evaluate the possibility of UPD for the chromosome of origin. In the same context, G-banded chromosome and FISH analysis may prove beneficial in cases with UPD by identifying or ruling out the presence of small marker chromosomes or mosaic trisomy. Furthermore, as shown in our study, a genotype-first approach to identification of recessive disorders is facilitated by identifying ROHs particularly in consanguineous families and those with increased risk for recessive Mendelian disorders. Selection of candidate genes for specific autosomal recessive disorders from the regions with allelic homozygosity, followed by sequencing analysis to unravel the homozygous mutations, reflects significant benefits of SNP array analysis in solving the puzzle for specific phenotypes.[1] Although it is challenging to limit the list of candidate disease genes to be considered in cases with extensive ROHs, analytical tools such as that developed by Wierenga *et al*[12] has proven invaluable especially when phenotypic information can be used to limit the genes/disorder to be considered. With the additional aid of whole genome or exon sequencing, this type of analysis may become the principal tool to finding molecular causes for disorders in patients with regions of allelic homozygosity.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

1 Sund KL, Zimmerman SL, Thomas C *et al*: Regions of homozygosity identified by SNP microarray analysis aid in the diagnosis of autosomal recessive disease and incidentally detect parental blood relationships. *Genet Med* 2013; **15**: 70–78.
2 Kearney HM, Kearney JB, Conlin LK: Diagnostic implications of excessive homozygosity detected by SNP-based microarrays: consanguinity, uniparental disomy, and recessive single-gene mutations. *Clin Lab Med* 2011; **31**: 595–613, ix.
3 Gibson J, Morton NE, Collins A: Extended tracts of homozygosity in outbred human populations. *Hum Mol Genet* 2006; **15**: 789–795.
4 Papenhausen P, Schwartz S, Risheg H *et al*: UPD detection using homozygosity profiling with a SNP genotyping microarray. *Am J Med Genet A* 2011; **155A**: 757–768.
5 Sasaki K, Mishima H, Miura K, Yoshiura K: Uniparental disomy analysis in trios using genome-wide SNP array and whole-genome sequencing data imply segmental uniparental isodisomy in general populations. *Gene* 2013; **512**: 267–274.
6 McQuillan R, Leutenegger AL, Abdel-Rahman R *et al*: Runs of homozygosity in European populations. *Am J Hum Genet* 2008; **83**: 359–372.
7 Campbell H, Carothers AD, Rudan I *et al*: Effects of genome-wide heterozygosity on a range of biomedically relevant human quantitative traits. *Hum Mol Genet* 2007; **16**: 233–241.
8 Rudan I, Smolej-Narancic N, Campbell H *et al*: Inbreeding and the genetic complexity of human hypertension. *Genetics* 2003; **163**: 1011–1021.
9 Ismail J, Jafar TH, Jafary FH, White F, Faruqui AM, Chaturvedi N: Risk factors for non-fatal myocardial infarction in young South Asian adults. *Heart (British Cardiac Society)* 2004; **90**: 259–263.
10 Grote L, Myers M, Lovell A, Saal H, Lipscomb Sund K: Variability in laboratory reporting practices for regions of homozygosity indicating parental relatedness as identified by SNP microarray testing. *Genet Med* 2012; **14**: 971–976.
11 Rehder CW, David KL, Hirsch B, Toriello HV, Wilson CM, Kearney HM: American College of Medical Genetics and Genomics: standards and guidelines for documenting suspected consanguinity as an incidental finding of genomic testing. *Genet Med* 2013; **15**: 150–152.
12 Wierenga KJ, Jiang Z, Yang AC, Mulvihill JJ, Tsinoremas NF: A clinical evaluation tool for SNP arrays, especially for autosomal recessive conditions in offspring of consanguineous parents. *Genet Med* 2013; **15**: 354–360.
13 Liehr T: Cytogenetic contribution to uniparental disomy (UPD). *Mol Cytogenet* 2010; **3**: 8.
14 Liehr T, Ewers E, Hamid AB *et al*: Small supernumerary marker chromosomes and uniparental disomy have a story to tell. *J Histochem Cytochem* 2011; **59**: 842–848.
15 Gao HZ, Kobayashi K, Tabata A *et al*: Identification of 16 novel mutations in the argininosuccinate synthetase gene and genotype-phenotype correlation in 38 classical citrullinemia patients. *Hum Mutat* 2003; **22**: 24–34.
16 Engel K, Hohne W, Haberle J: Mutations and polymorphisms in the human argininosuccinate synthetase (ASS1) gene. *Hum Mutat* 2009; **30**: 300–307.
17 Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, Wilson JF: Genomic runs of homozygosity record population history and consanguinity. *PLoS One* 2010; **5**: e13996.
18 Grote L, Myers M, Lovell A, Saal H, Sund KL: Variable approaches to genetic counseling for microarray regions of homozygosity associated with parental relatedness. *Am J Med Genet A* 2013; **164A**: 87–98.
19 Schaaf CP, Scott DA, Wiszniewska J, Beaudet AL: Identification of incestuous parental relationships by SNP-based DNA microarrays. *Lancet* 2011; **377**: 555–556.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)